Delft University of Technology

Application of GIS-based machine learning algorithms for prediction of irrigational groundwater quality indices

Mohammed, Musaab A. A.; Kaya, Fuat; Mohamed, Ahmed; Alarifi, Saad S.; Abdelrady, Ahmed ; Keshavarzi, Ali; Szabó, Norbert P.; Szűcs, Péter

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Check for updates

Application of GIS-based machine learning algorithms for prediction of irrigational groundwater quality indices

Musaab A. A. Mohammed[1,2]*, Fuat Kaya[3], Ahmed Mohamed[4], Saad S. Alarifi[5], Ahmed Abdelrady[6]*, Ali Keshavarzi[7], Norbert P. Szabó[1,8] and Péter Szűcs[1,8]

[1]Faculty of Earth Science and Engineering, University of Miskolc, Miskolc, Hungary, [2]College of Petroleum Geology and Minerals, University of Bahri, Khartoum, Sudan, [3]Faculty of Agriculture, Isparta University of Applied Sciences, Isparta, Türkiye, [4]Geology Department, Faculty of Science, Assiut University, Assiut, Egypt, [5]Department of Geology and Geophysics, College of Science, King Saud University, Riyadh, Saudi Arabia, [6]Faculty of Civil Engineering and Geoscience, Delft University of Technology, Delft, Netherlands, [7]Laboratory of Remote Sensing and GIS, University of Tehran, Karaj, Iran, [8]MTA-ME Geoengineering Research Group, University of Miskolc, Miskolc, Hungary

Agriculture is considered one of the primary elements for socioeconomic stability in most parts of Sudan. Consequently, the irrigation water should be properly managed to achieve sustainable crop yield and soil fertility. This research aims to predict the irrigation indices of sodium adsorption ratio (SAR), sodium percentage (Na%), permeability index (PI), and potential salinity (PS) using innovative machine learning (ML) techniques, including K-nearest neighbor (KNN), random forest (RF), support vector regression (SVR), and Gaussian process regression (GPR). Thirty-seven groundwater samples are collected and analyzed for twelve physiochemical parameters (TDS, pH, EC, TH, Ca^{+2}, Mg^{+2}, Na^+, HCO_3^-, Cl, SO_4^{-2}, and NO_3^-) to assess the hydrochemical characteristics of groundwater and its suitability for irrigation purposes. The primary investigation indicated that the samples are dominated by Ca-Mg-HCO_3 and Na-HCO_3 water types resulted from groundwater recharge and ion exchange reactions. The observed irrigation indices of SAR, Na%, PI, and PS showed average values of 7, 42.5%, 64.7%, and 0.5, respectively. The ML modeling is based on the ion's concentration as input and the observed values of the indices as output. The data is divided into two sets for training (70%) and validation (30%), and the models are validated using a 10-fold cross-validation technique. The models are tested with three statistical criteria, including mean square error (MSE), root means square error (RMSE), and correlation coefficient (R^2). The SVR algorithm showed the best performance in predicting the irrigation indices, with the lowest RMSE value of 1.45 for SAR. The RMSE values for the other indices, Na%, PI, and PS, were 6.70, 7.10, and 0.55, respectively. The models were applied to digital predictive data in the Nile River area of Khartoum state, and the uncertainty of the maps was estimated by running the models 10 times iteratively. The standard deviation maps were generated to assess the model's sensitivity to the data, and the uncertainty of the model can be used to identify areas where a denser sampling is needed to improve the accuracy of the irrigation indices estimates.

1 Introduction

Groundwater is one of the primary providers of irrigation water for agriculture, offering a dependable and sustainable source of irrigation water (Li et al., 2022). The utilization of groundwater for irrigation offers various benefits, including dependability and regularity. Unlike surface water, which can be impacted by floods and droughts, groundwater is generally consistent and can ensure a continuous supply of irrigation water (Mohammed et al., 2022c). This is crucial in areas with insufficient or unstable surface water supplies (Osta et al., 2022). Additionally, groundwater can be used in combination with surface water to create a more dependable and long-lasting source of irrigation water (Mohamed et al., 2023). Farmers can lessen their reliance on surface water and improve the effectiveness of their irrigation systems by employing both surface water and groundwater. However, aquifers need to be managed and protected to ensure the long-term sustainability of groundwater as a source of irrigation water (Docheshmeh Gorgij et al., 2022). The eastern Nile River region is mostly dependent on groundwater for irrigation because it is agricultural terrain (Mohammed et al., 2023d). The high reliance on groundwater for irrigation is due to the absence of surface water transporting systems and the high cost of delivering Nile water to agricultural lands (Farah et al., 2000). As a result, and due to the expanding agricultural lands and over-pumping of groundwater aquifers, the groundwater quality for domestic and agricultural purposes is declining (Eyankware et al., 2022). The quality of groundwater for irrigation purposes is crucial since it determines crop yields, soil fertility and permeability (El Bilali et al., 2021). In order to produce crops sustainably, this project aims to assess the suitability of groundwater for irrigation using advanced computational artificial intelligence (AI) systems.

Irrigational water quality might vary greatly based on the quantity and quality of soluble salts. Other parameters, such as variations in climatic conditions, may significantly influence the irrigational groundwater (Sattari et al., 2018). For instance, groundwater quality is poorer during the dry seasons compared to rainy seasons due to delusion by groundwater recharge. Therefore, water quality evaluations should be made during each irrigation cycle. Some physiochemical parameters are present in all irrigation water, which can have major impacts even with tiny concentrations. The type of salt present in the water and its overall concentration affects the suitability of the irrigation water. Salt can harm soil and plants regardless of its form or quantity. Consequently, several water quality indices were proposed considering different physiochemical parameters on groundwater or the impact of the irrigational water on the soil (Wilcox, 1948; Richards, 1954; Doneen, 1964; Kelly and Reiter, 1984). For instance, sodium percentage (Na%) and sodium adsorption ratio (SAR) determine the amount of sodium relative to the other cations and thus detect the effect of the cation exchange process on soil permeability (Chidambaram et al., 2022). Permeability index (PI) directly measures the influence of a particular water quality type on the soil (Kouadra and Demdoum, 2020), while potential salinity (PS) indicates the suitability of irrigational water for particular plants (Masoud et al., 2022). Usually, in assessing irrigational water quality, these parameters are combined for comprehensive evaluation. Irrigation indices are effective instruments for determining whether groundwater is suitable for agricultural use. However, the calculation of these indices is often lengthy and time-consuming (Nouraki et al., 2021); therefore, AI techniques are proposed to reduce the calculation time and avoid calculation errors (Nabiollahi et al., 2021). Because AI models can evaluate vast amounts of data and produce precise forecasts, their usage in irrigation water management has increased recently (Yu et al., 2022; Masoudi et al., 2023).

Groundwater and/or water quality studies have extensively used AI and machine learning (ML) approaches (Ahmed et al., 2019; Abdel-Fattah et al., 2021; Mohammed et al., 2022b, 2022a; Docheshmeh Gorgij et al., 2022; Najafzadeh et al., 2022; Nasir et al., 2022; Nong et al., 2023). This paper aims to introduce the application of AI techniques in irrigation water management, with a focus on the prediction of irrigation water indices. Recently, the use of AI techniques for the prediction of irrigation indices has been growing due to the high non-linearity and complexity of these indices (Pipia et al., 2019; Rahnama et al., 2020; Sattari et al., 2020; Dimple et al., 2022; Trabelsi and Ali, 2022). For instance, Mokhtar et al. (2022) used support vector regression (SVR) and random forest (RF) to model the irrigation water quality of potential salinity, sodium percentage and permeability index in Bahr El-Baqr, Egypt. They indicated the robustness of these algorithms to support the decision-making process for sustainable crop yield. Yahyaoui et al. (2023) conducted a comparative study to examine the capabilities of several ML algorithms, including K-nearest neighbor (KNN), support vector machine (SVM) and decision trees in predicting irrigational water quality indices (IWQI) in Cap-Bon, Tunisia. Their study revealed the efficiency of KNN techniques over the others. Singh (2020) employed SVM and Gaussian process regression (GPR) for the simulation of SAR in three sub-watersheds in Iran. These studies demonstrated the potential of AI and ML as a tool for predicting various water quality indices in irrigation systems and highlighted the importance of such predictions in improving water management practices and ensuring sustainable agriculture. However, these studies rely only on the prediction of the observed values without examining the capabilities of AI techniques in detecting the spatial distribution of the predicted indices. In this research, we proposed a geospatial prediction methodology based on integrating geographical information systems (GIS) with ML algorithms. Other techniques, such as Remote sensing and geostatistics, are widely used for monitoring urban water supplies and assessing their potential for future advancements (Liu et al., 2023). In addition, an adaptive model was constructed for the purpose of correcting water depth bias correction (Zhou et al., 2023).

This paper aims to examine the capability of several geospatial AI algorithms, including KNN, RF, SVR, and GPR integrated with GIS to predict the spatial distribution of irrigational water quality indices of Na %, SAR, PS and PI. The results of this research improve irrigation water management and the efficiency and sustainability of agricultural production.
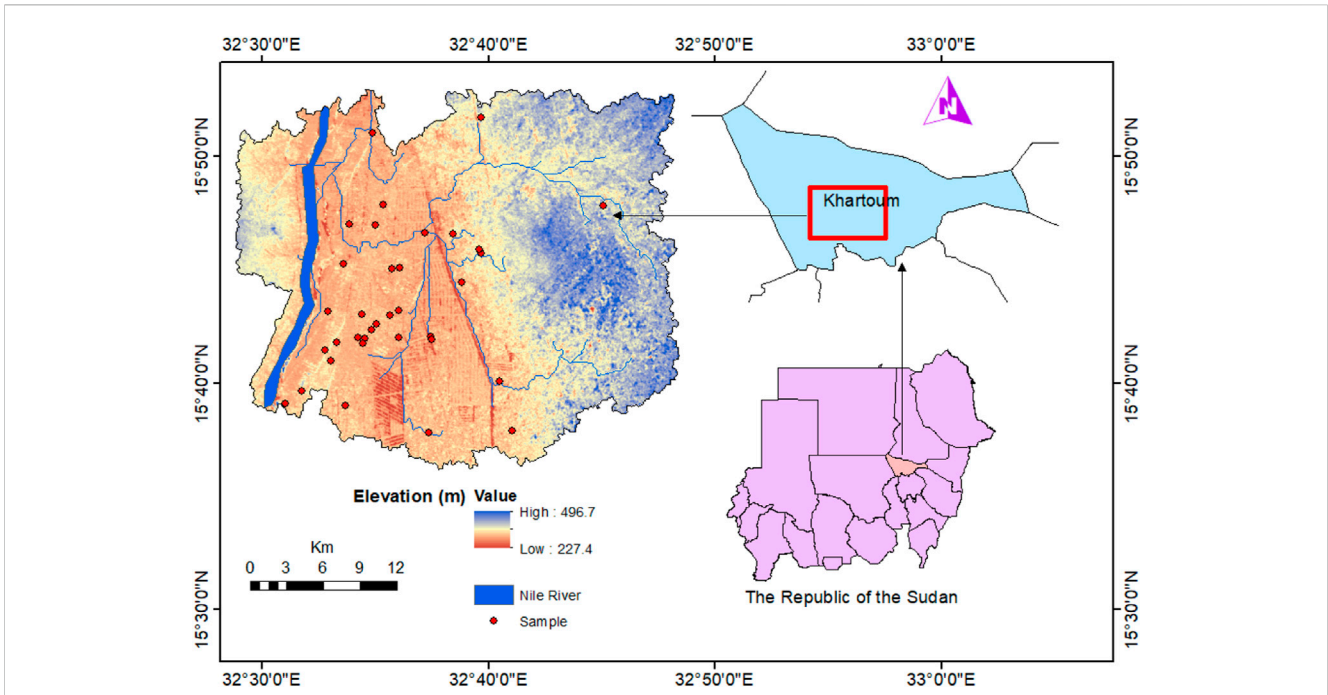
FIGURE 1
The primary geomorphological features are depicted by the research area's location.
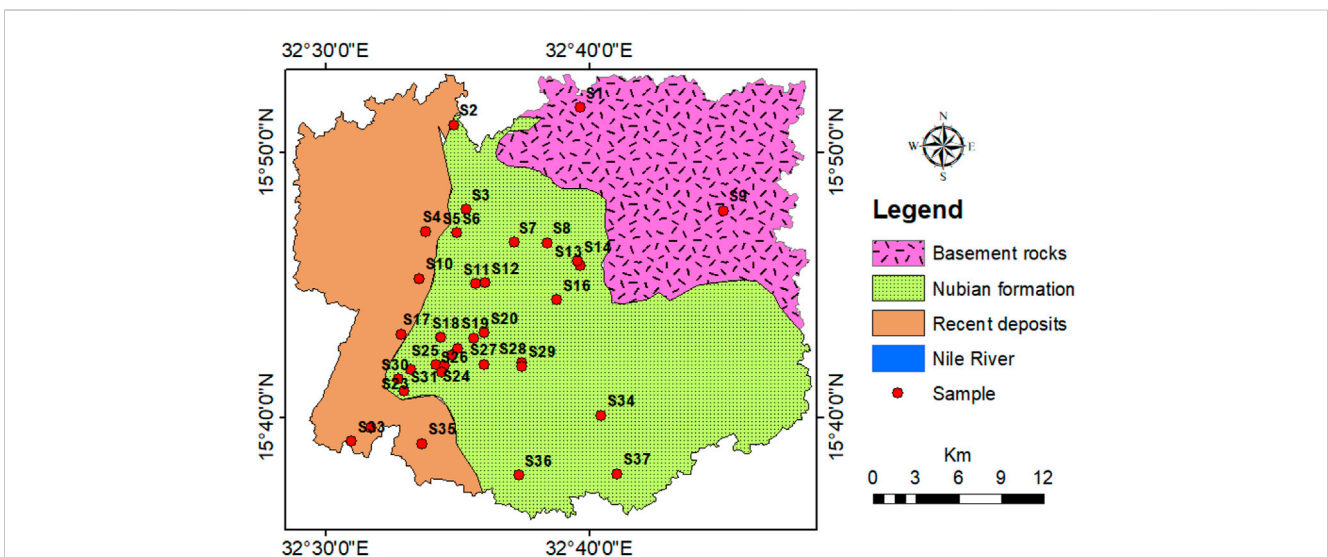
FIGURE 2
Geological map illustrates the main geological units in the study area.

2 Materials and methods

2.1 Study area

This study explores the suitability of groundwater for irrigation purposes in the eastern Nile River area, Khartoum state, Sudan. The study area lies between longitude 32° 30′ and 32° 47′ and latitude 15° 34′ and 15° 55′ (Figure 1). The area is characterized by a hot climate in summer, cold and dry in winter, and associated with an annual average precipitation of 115.7 mm/year in the fall season. The main geomorphological features are the Nile and Blue Nile Rivers, which bound the study area from the west. In general, the topography of the study area is flat, with an elevation range from 227 m above mean sea level (a.m.s.l) in the western and central parts to more than 497 m in the eastern part of the area (Figure 1).

Geologically, the area is located in the Blue Nile rift basin, where three geological units dominate. Figure 2 illustrates the primary geological units observed in the study area. The Pan African basement rocks of the

FIGURE 3
The groundwater flow map shows the areal variation in groundwater levels.
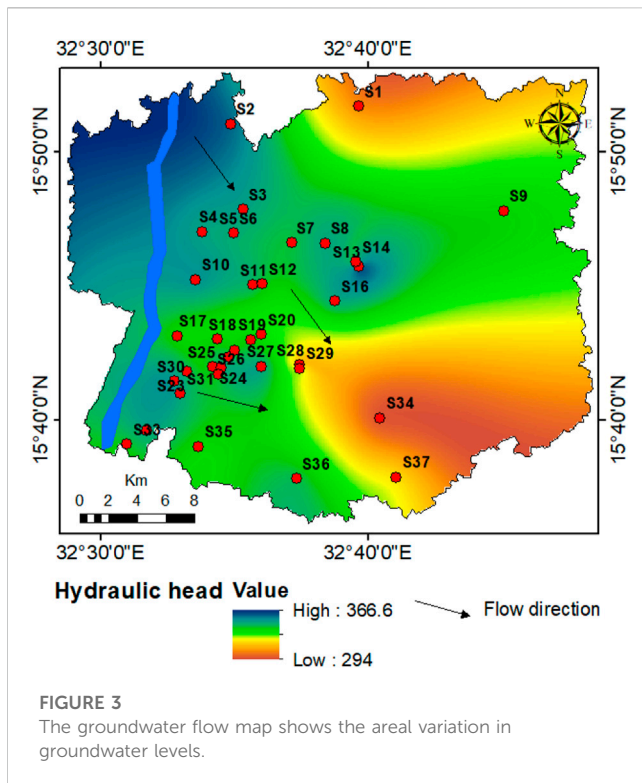
Precambrian age form the bottom of the Blue Nile basin (Idriss et al., 2011; Mohammed et al., 2023e). These rocks are dominated by biotite granite, gneiss, and schist mainly observed near Khartoum's northern and eastern boundaries (Awad, 1994). The Precambrian basement rocks are overlain by mudstone, sandy mudstone, conglomerates, and sandstone, which have been consolidated by limestone, siliceous, and ferrous minerals (Mohammed et al., 2023a). This rock accumulation is known as Cretaceous Nubian Formation (Mohammed et al., 2023f). This formation also comprises evaporite deposits formed in a braided environment and dispersed throughout the Nile and Blue Nile Rivers (Hussein, 1992). The recent deposit of Quaternary age is observed in the surroundings of the Blue and Nile Rivers and the eastern part of the study area. This geological unit is also known as the Gezira formation and comprises unconsolidated sand, gravel, and silts (Whiteman, 1971).

The Nubian sandstone, with an average thickness of 300 m, serves as a primary groundwater aquifer in the study area (Köhnke et al., 2017). This aquifer is classified as highly productive, with an average transmissivity of 700 m²/day (Elkrail and Adlan, 2019; Mohammed M. et al., 2023). Given that there is relatively minimal recharge from rainfall, the Nile River and ephemeral streams are the primary sources of groundwater recharge to the Nubian formation (Mohammed et al., 2023c). As a result, groundwater levels range from 366.6 m in the western parts near Nile River to 294 m in the eastern part (Figure 3). Consequently, groundwater flows mainly from the western to the eastern part of the region.

2.2 Groundwater sampling

As part of the "zero thirsty" program administered by the Sudanese government, the Khartoum State Water Corporation

collected 37 groundwater samples in December 2020. The aim of the zero thirsty project was to avail a clean and sustainable water supply in the Sudanese states. Khartoum state, as the most populated, was part of this project in which the suitability of groundwater for drinking purposes needs to be determined. The groundwater samples were collected from public and privately owned groundwater wells with a depth ranging from 100 to 250 m (Mohammed et al., 2023f). The sampling protocol is followed during the data collection. The groundwater wells are pumped for 30 min before the sample collection, and the samples are preserved in a polyethylene bottle washed with deionized water. The containers were transferred to the lab in an ice-filled box after being packed securely. The groundwater samples were analyzed in the labs of Groundwater and Wadies Directorate for eleven (11) physiochemical parameters. The parameters include total dissolved solids (TDS), hydrogen ion activity (pH), electrical conductivity (EC), total hardness (TH), calcium (Ca), magnesium (Mg), sodium (Na), bicarbonate (HCO_3), chloride (Cl), sulfate (SO_4), and nitrate (NO_3). EC, TDS, and pH was determined shortly after sample collection using a portable multi-parameter equipment. While for Cl, SO_4, NO_3, Ca, Mg, Na, ion chromatography (IC) is used to analyze these parameters by separating ions based on their charge and affinity to the stationary phase.

Since groundwater fulfills the principle of electrical equilibrium, the electrical balance (EB) between the major cations and anions in meq/L is calculated using Appelo and Postma (2005) formula (Eq. 1) to examine the accuracy of the hydrochemical analysis. In this investigation, the EB for all the analyzed samples ranged between +10 and −10, which indicates acceptable accuracy.

$$(EB\%) = \frac{\sum cations - \sum anions}{\sum cations + \sum anions} \times 100 \qquad (1)$$

2.3 Irrigation indices

The quality of irrigation water and its suitability for various crops are assessed using irrigation water quality indices (Mallik et al., 2022). There are several different irrigation water indicators; however, in this research, four indices are used for the management of irrigation water, including sodium adsorption ratio (SAR), sodium percentage (Na%), permeability index (PI), and potential salinity (PS). The sodium adsorption ratio (SAR) indicates the amount of sodium in water and how it could impact crops and soil (Richards, 1954). Low SAR levels are often regarded as acceptable for irrigation, whereas high SAR values can cause soil dispersion and poor crop development. The overall amount of sodium in the water is also determined by the sodium percentage (Na %) (Wilcox, 1948). The capacity of irrigation water to infiltrate into the soil is measured by the permeability index (PI), which highly influences plant growth by affecting the absorption of water by plants (Falowo et al., 2017). The total amount of dissolved salts in the water is measured by potential salinity (PS) (Doneen, 1964). High PS values can cause the soil to become salinized, which can harm crop development. The following formulas with the physiochemical parameters measured in meq/L can be used to determine the irrigation indices.

TABLE 1 Machine learning algorithm parameters used to model irrigation indices.

Algorithms	Irrigation indices	Tuning hyperparameter
K nearest neighbors (KNN)	SAR	k = 2
	Na %	k = 8
	PI	k = 4
	PS	k = 8
Gaussian process regression (GPR)	SAR	sigma = 0.061
	Na %	sigma = 0.049
	PI	sigma = 0.395
	PS	sigma = 0.119
Random forest (RF)	SAR	mtry = 6
	Na %	mtry = 6
	PI	mtry = 6
	PS	mtry = 6
Support vector regression (SVR)	SAR	epsilon = 0.1, sigma = 0.003, cost (C) = 18.86
	Na %	epsilon = 0.1, sigma = 0.07, cost (C) = 18.27
	PI	epsilon = 0.1, sigma = 0.015, cost (C) = 767.97
	PS	epsilon = 0.1, sigma = 0.07, cost (C) = 539.10

$$SAR = \frac{Na^+}{\sqrt{\frac{Ca^{+2}+Mg^{+2}}{2}}} \qquad (2)$$

$$Na^+\% = \frac{Na^+}{Ca^{+2} + Mg^{+2} + Na^+} * 100 \qquad (3)$$

$$PI = \frac{Na^+ + \sqrt{HCO_3^-}}{Ca^{+2} + Mg^{+2} + Na^+} * 100 \qquad (4)$$

$$PS = Cl^- - (SO_4^{-2}/2) \qquad (5)$$

2.4 Machine learning models

Four machine learning models were used in this study to predict irrigation indices: KNN, GPR, SVR, and RF. The dataset was separated into two parts: 70% for calibrating the machine learning models and 30% for validating the models. The analysis of the ML algorithms was conducted using the R Core Environment version 4.2.1 (Team, 2009). Evaluations specific to each algorithm are presented in separate sections, and the parameters used in the model application process are presented in tables in the following part.

2.4.1 K-nearest neighbors (KNN)

The KNN tries to identify the k-nearest instances in the training set and assigns the instance to the label that occurs most frequently within that k-subset. If the values are continuous, the target can be calculated by computing the mean. Nearest-neighbor methods utilize the observations in the calibration set that is closest in input space to x to form Y (Hastie et al., 2009). The k-nearest neighbor fit for Y is specifically defined as

$$Y(x) = \frac{1}{k} \sum_{x_i \in N_k(x)} yi, \qquad (6)$$

where $N_k(x)$ is the neighborhood of x, defined as the k closest points x_i in the training sample. The concept of closeness is predicated on a metric, which is considered to be the Euclidean distance in this example. As a result, we locate the k observations closest to x in input space and average their responses. For a more in-depth explanation of the KNN method, please refer to Hastie et al. (2009). In this study, the "train" function in the "caret" package (Kuhn et al., 2020) of the R Core Environment software (Version 4.2.1) was utilized to implement the KNN algorithm. The optimal number of k was determined through a parameter searching process, and the results for the irrigation indices models are presented in Table 1.

2.4.2 Random forest (RF)

The RF (Breiman, 2001) is an enhanced version of bagging that involves constructing a large number of uncorrelated trees and combining their predictions. For a more detailed understanding of the Random Forest method, one can consult Biau and Scornet, (2016). The "randomForest" package by Liaw et al. (2002) was used in this study. The optimal value of the parameter "*mtry*" was determined through a parameter tuning process using the "train" function in the "caret" package in the R Core Environment software (Version 4.2.1). In this study, the default settings were used for the other parameters (etc., ntree) in the "randomForest" package. The optimal number of *mtry* was determined through a parameter searching process and the results for the irrigation indices models are presented in Table 1.

2.4.3 Gaussian process regression (GPR)

In this study, GPR (Rasmussen et al., 2006) was employed to examine the relationship between water quality parameters and spatial distribution of irrigation indices. GPR was used for both inference and mapping purposes. GPR aims to reconstruct the underlying signal f by removing the contaminating noise ε. For a deeper understanding of the GPR method, can refer to the publication by Rasmussen et al. (2006) and Ballabio et al. (2019). The Gaussian radial basis function (RBF) kernel is one of the most popular kernels used in Gaussian Process Regression, and it can model non-linear relationships between the input variables and the target variable. By using the "gaussprRadial" method in the "train" function of the "caret" package, the GPR algorithm was implemented using the Gaussian RBF kernel in the R Core Environment software (Version 4.2.1). The parameter of the GPR models were optimized by using a repeated 10-fold cross-validation method to prevent overfitting and presented in Table 1.

2.4.4 Support vector regression (SVR)

The SVR is a machine-learning technique that utilizes kernels to map the input space to a high-dimensional feature space, allowing for non-linear mapping (Drucker et al., 1996). The goal of SVR is to reduce both prediction errors and model complexity simultaneously. The optimization problem is solved using Lagrange multipliers and results in a set of support vectors that define the boundary. The prediction for a new data point is then made based on the support vectors and their weights. In this study, the SVR method was implemented using the "e1071" package (Meyer et al., 2020) in the R Core Environment software (Version 4.2.1). The model's parameters, including the choice of kernel function and the value of parameter c, were determined through a parameter tuning process using a random search method. The optimal values for these parameters are reported in Table 1. For a more detailed understanding of the SVR method, one can consult Hastie et al. (2009).

2.4.5 Parameter optimization of the models

The KNN, GPR, SVR, and RF algorithms require the setting of several user-defined hyper-parameters. To find the optimal values for these parameters, the "caret" package (Kuhn et al., 2020) via R software (Team, 2009) was employed in this study. This tuning process computed a 10-fold cross-validation and random search method to determine the optimal values (Table 1). Subsequently, Kriging method is used to visualize and model the predicted parameters. It offers a valuable tool for generating spatially continuous maps that accurately represent the distribution of a irrigational indices across an area. It is particularly effective when data is sparse or irregularly distributed and when capturing spatial trends and correlations is crucial for analysis.

2.5 Models performance and spatial uncertainty

In this study, the accuracy of the predicted and observed data was evaluated using three widely metrics: root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). To accurately reflect the overall accuracy of the machine learning models, only the evaluation metrics for the validation set are calculated.

$$MAE = \frac{\sum_{i=1}^{n} |Oi - Pi|}{n} \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Oi - Pi)2}{n}} \quad (8)$$

$$R^2 = \left[\frac{\sum_{i=1}^{n} (Oi - Oave)(Pi - Pave)}{\sqrt{\sum_{i=1}^{n} (Oi - Oave)^2 (Pi - Pave)^2}} \right]^2 \quad (9)$$

where O_i and P_i are, respectively, the observed and predicted values, with their average values represented by O_{ave} and P_{ave}, respectively, and n is the sample size in validation set. Most previous studies have aimed to enhance the performance of data-based models in estimating irrigation indices. However, the overall spatial uncertainty of the models used for modeling irrigation indices has not been assessed yet. Therefore, uncertainty was calculated as the standard deviation of the predictions of irrigation indices in each pixel over 10 iterations, which represents the spread of predicted values around the mean in each pixel (Yigini et al., 2018).

3 Results and discussion

3.1 Hydrochemical investigation

A number of eleven parameters are used to investigate the hydrochemical characteristics of the groundwater samples. The concentration of the parameters resulting from the hydrochemical analysis is illustrated in Table 2, while Table 3 summarizes the descriptive statistics of the analyzed parameters. The comparison of between the concentration of the parameters and the standard of the World Health Organization (WHO) (Edition, 2011) indicated that the maximum concentration of most parameters exceeded the permissible limit. The amount of the TDS ranged from 190.2 to 1742 mg/L, with the highest concentration observed in S27. TDS is considered the most important parameter in studying groundwater salinity (Freeze and Cherry, 1979; Mohammed et al., 2022c). Since groundwater with TDS higher than 1,000 mg/L is considered as brackish water (Carroll, 1962), 8% of the samples were classified as brackish, while the remaining percentage is considered freshwater. The pH of groundwater samples varied between 6.5 and 8.59 with a mean value of 7.6, denoting that most of the groundwater samples are neutral to alkaline, with one sample (S17) above the acceptable limit of WHO (Edition, 2011). The range of the EC content in groundwater samples is from 317 to 2,620 μS/cm. Three samples, including S2, S6, and S27, exceeded the EC standard limit of 1,500 μS/cm. The concentration of TH ranged from 124 to 890 mg/L. The maximum concentration is recorded in S6, and the minimum is in S29. According to Sawyer and McCarty (1967) classification, 78.3% of the groundwater samples are considered hard water (TH 150–300 mg/L), while the remaining 21.6% are very hard water (TH > 300). With an average value of 47 mg/L, the Ca^{+2} concentration ranged from 11 to 101 mg/L. while for Mg^{+2} concentrations ranged from 5.8 mg/L in the S12 location to 82.6 mg/L in the S16. Na^+ concentration varied between 10 and

TABLE 2 The findings derived from the hydrochemical study conducted on the gathered groundwater samples.

Sample no.	TDS (mg/L)	pH	EC (μS/cm)	TH (mg/L)	Ca^{+2} (mg/L)	Mg^{+2} (mg/L)	Na$^+$ (mg/L)	HCO$_3^-$ (mg/L)	Cl$^-$ (mg/L)	SO$_4^{-2}$ (mg/L)	NO$_3^-$ (mg/L)
S1	329	7.4	479	269	65	23	10	260	12.8	33	8.35
S2	988	7.8	1,648	316	53.6	43.7	137.9	382	50	174	3
S3	427	7.5	618	166	11	33.5	144.7	451.4	11.4	46	0.88
S4	362.6	7.4	518	185	38.4	22.4	45.9	281.6	23.9	23	6.16
S5	626	7.5	1,253	545	87.5	78.3	138.8	500	60	227	4.4
S6	1,267	7.2	1,810	890	52	51	161	620	86.6	89.7	0.04
S7	959	7.5	1,370	310	40	21	115	335	119	174	24.2
S8	694.4	7.7	929	260	57.6	28.2	123	292.8	91.5	152	7.48
S9	303	7	552	220	60	16.3	25.8	180	36	38	14.5
S10	330	7.7	620	224	60	29	45	270	16	30	0.0012
S11	359	7.4	598	148	30.4	17.3	74.7	212	24	53	3.5
S12	600.6	7.8	858	190	66.4	5.8	124.3	378.2	82.5	86	5.72
S13	411.6	7.5	588	230.6	48.64	26.5	88.4	396.5	11.4	37	13.64
S14	724.9	7.7	1,318	300	44	45.6	147.5	340	120	100	2.8
S15	483	7.4	690	232	27.2	39.8	67.9	219	61	112	14.5
S16	1,050	7.9	1,500	380	16	82.6	332.8	414.8	193	320.5	0.07
S17	452.2	8.59	646	256	54.4	29.2	54.9	280.6	16.3	38	1.32
S18	438.2	7.14	626	226	20.8	42.3	36.1	353.8	35.5	37	0.88
S19	375	7.5	650	232	52.8	24	46.5	276	18	41	1.4
S20	331	8.5	474	192	48	17.5	26.2	231	11.4	35	3.52
S21	354	7.4	590	208	40.3	25.4	37.2	248	8	28	1.2
S22	322	7.2	585	200	38.4	25	46.8	250	12	27.9	3
S23	339.5	7.4	485	254	56	27.7	37	414.8	12.8	98	9.24
S24	326	7.6	593	216	48	23	48.4	280	12	34	1
S25	330.8	7.3	601.5	234	35.2	35	37.6	226	23	52	3.2
S26	483	7.4	690	232	27.2	39.8	67.9	219	61	112	14.5
S27	1,742	8.3	2,620	404	75.2	52.5	640	414.8	153.4	650	9.24
S28	214.4	8	357.3	134	24	17.8	19.8	142	9	14	5.3
S29	190.2	7.9	317	124	27.2	13.44	24.1	130	20	13	3.7
S30	292	8.3	417	186	32	25.8	17.6	231.8	120.7	4	0.1
S31	222	7.4	370	160	39.2	14.88	14.4	186	4	3	7.3
S32	265	7.4	540	150	35	36	60	260	62	50	0.003
S33	569	7.7	1,036	400	101	35	32.4	235	136	58.8	4.5
S34	529	7.9	757	319	94.9	19.9	107.7	416	35.5	51	70
S35	413	7.2	590	186	32	21.4	78.1	317	25.6	47	3.96
S36	590	7.58	1,071	264	41.6	38.4	114.3	326	44	120	6.16
S37	304.9	6.5	554.4	156	38.4	14.4	41.9	98	80	33	10.5

TABLE 3 The descriptive statistics of the measured parameters compared to WHO guidelines (Edition, 2011).

Parameter	Unit	Min	Mean	Maximum	WHO guidelines
TDS	mg/L	190.2	536.7	1742	1,000
pH	—	6.5	7.6	8.59	6.5–8.5
EC	µS/cm	317	842	2,620	1,500
TH	mg/L	124	272	890	500
Ca^{+2}	mg/L	11	47	101	200
Mg^{+2}	mg/L	5.8	31.5	82.6	150
Na+	mg/L	10	103	640	200
HCO_3^-	mg/L	98	302	620	350
Cl^-	mg/L	4	53.7	193	250
SO_4^{-2}	mg/L	3	99.9	650	200
NO_3^-	mg/L	0.0012	8.7	70	50

640 mg/L, with S1 showing the lowest concentration and S27 the highest. Only 5.4% (S16 and S27) exceeded the WHO-allowable level (Edition, 2011). The HCO_3^- concentration ranges from 98 to 620 mg/L, with an average of 302. Eleven samples had HCO_3^- concentration higher than the WHO recommendations. Since the Cl^- concentration ranges from 4 to 193 mg/L, all groundwater samples exhibit Cl^- contents below guidelines. The concentration of SO_4^{-2} in groundwater samples ranges from 3 mg/L in S31 to 650 mg/L in S27. Samples in the study area had NO_3^- contents ranging from 0.0012 to 70 mg/L. S34 shows the highest concentration, whereas S10 shows the lowest. Only one sample (S34) had NO_3^- levels that were above WHO standards.

The regional distribution of these factors is depicted in Figure 4. The majority of the parameters demonstrate a consistent pattern as the concentration progressively increases from the western to the eastern region of the research area. Unlike most of the parameters HCO_3^- and Ca^{+2}, the most significant concentration is situated in the western region of the research area, which is likely an indication of the impact of Nile water infiltration, which is likely to be highly concentrated with HCO_3^- and Ca^{+2} (Mohammed et al., 2022b). The change in the hydrochemical facies is also studied with the aid of Chadha (1999) diagram. In this diagram, the difference between major cations ($Ca^{+2} + Mg^{+2}$) - ($Na^+ + K^+$) and anions [HCO_3 - (SO_4^{-2} + Cl^-)] is used to detect the groundwater types (Figure 5). Consequently, four groundwater facies are revealed as Na-Cl, Ca-Mg-SO_4/Cl, Na-HCO_3, and Ca-Mg-HCO_3. Most of the groundwater samples (67.5%) fall in Ca-Mg-HCO_3 water type, which indicates the influence of groundwater recharge on groundwater chemistry. The locations of these samples are within the influence of the Nile River, which is 12 km (Farah et al., 2000). The groundwater type gradually changes from the western to the eastern parts of the study area from Ca-Mg-HCO_3 to Na-HCO_3 water type. This change is likely due to ion exchange or the replacement of Ca^{+2} and Mg^{+2} with Na^+. As a result, 16.2% of groundwater samples are identified as Na-HCO_3 water type. 8.1% of the samples are classified as Ca-Mg-SO_4/Cl resulting from reverse anion exchange in which HCO_3^- is replaced by Cl^- in groundwater. The continuation of cation and reverse anion exchange leads to the Na-Cl facies (Abdelsalam et al.,

2016). In this study, 8.1% are classified as saline water. The hydrochemical attributes of groundwater in the eastern Nile River region are predominantly governed by the processes of groundwater recharge and ion exchanges.

3.2 Irrigation indices

Agricultural activities are one of the primary proficient in central Sudan, and groundwater is considered a primary source for irrigation. The quality of groundwater used for irrigation is highly influencing soil fertility and crop growth; as a result, the suitability of irrigational water must be assessed. In this research, four indices are used to evaluate the quality of groundwater for irrigation, including SAR, Na^+ %, PI and PS. The estimated indices of the groundwater samples utilized in this study are presented in Table 4.

3.2.1 SAR

The SAR ranged from 0.27 to 13.8. The classification of groundwater samples is represented in Figure 6A (USSL diagram). Groundwater is divided into four groups based on SAR: excellent with SAR < 10 (S1); good [SAR ranges from 10 to 18 (S2)]; doubtful, in which SAR ranges from 18 to 26 (S3); and unsuitable with SAR > 26 (S4) (Richards, 1954). In general, SAR is influenced by the concentration of Na^+ relative to the other cations such as Ca^{+2}, Mg^{+2} and K^+ (Rawat et al., 2018). In practice, groundwater is usually classified by conjugation of SAR with EC, since irrigation water with high salinity stimulates the ion exchange process and thus affects the adsorption of water by plants. Salinity. Based on EC, groundwater is classified as water with low (C1), medium (C2), high (C3), and very high (C4) salinity hazard. As a result, 59.4% of the groundwater samples are associated with low SAR (S1) and medium salinity hazard (C2). This class is considered excellent for irrigation purposes. 35% of the samples are projected in S1C3 class with low alkali and high salinity hazard. This class might not affect the soil permeability however, high salinity may influence the growth of salinity-sensitive plants and thus reduces the crops yield. One sample (S16) is plotted in S2C3 class with medium alkali and high
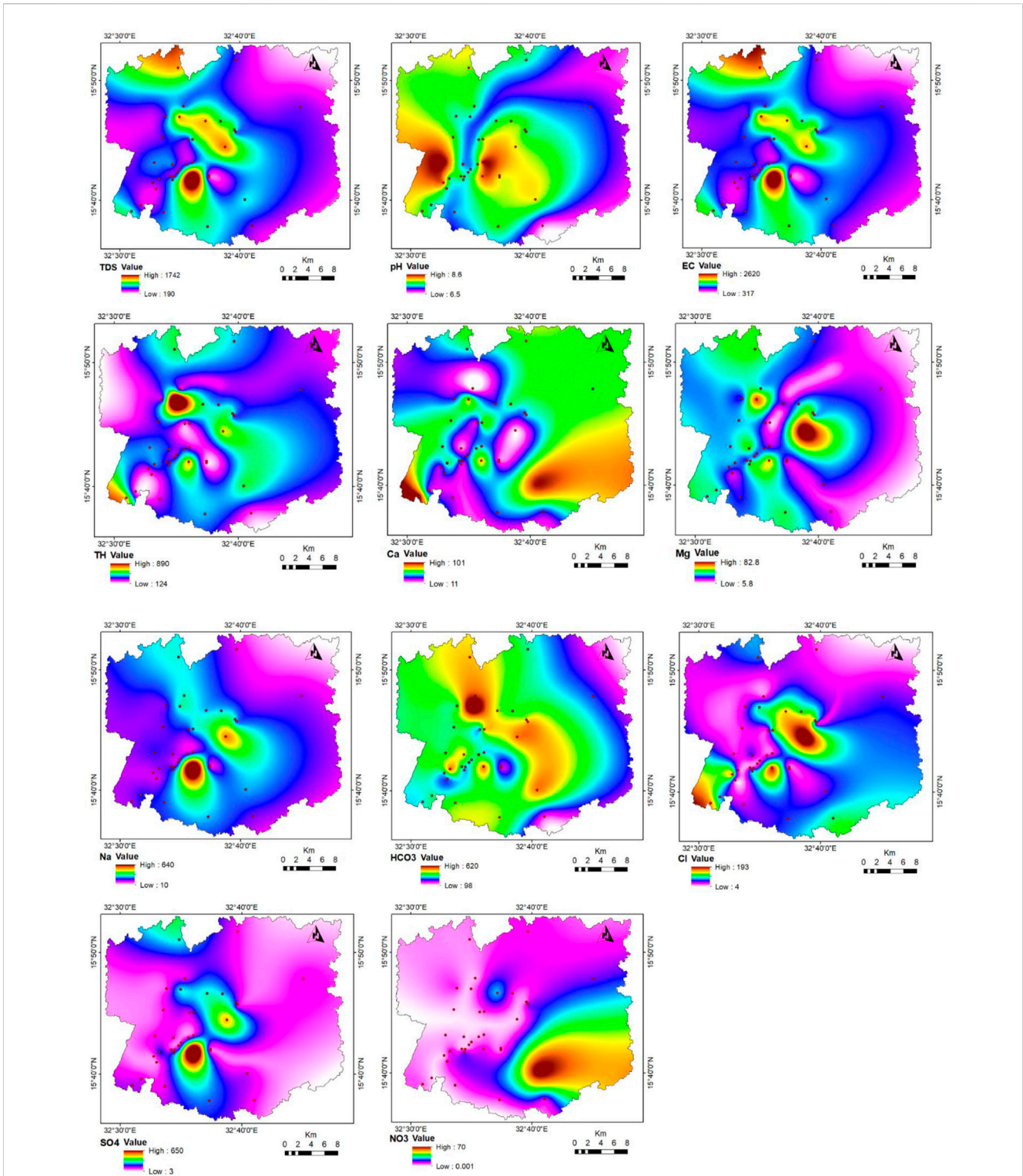
FIGURE 4
The areal variation of the analyzed parameters used for groundwater quality evaluation.

salinity hazards. S27 is classified as unsuitable for irrigation since it is associated with high alkali and high salinity hazards (S4C4). This persistent use of this sample for irrigation will damage the soil permeability by incorporating Na^+ within the soil particles and affect the growth of plants.

3.2.2 Na^+%

The principle of Na% is almost similar to that of SAR in which the percentage of Na^+ relative to the cations of Ca^{+2} and Mg^{+2} is measured. Na^+ is incorporated into the clay minerals sheets while the other cations are removed, which affects the infiltration of water to
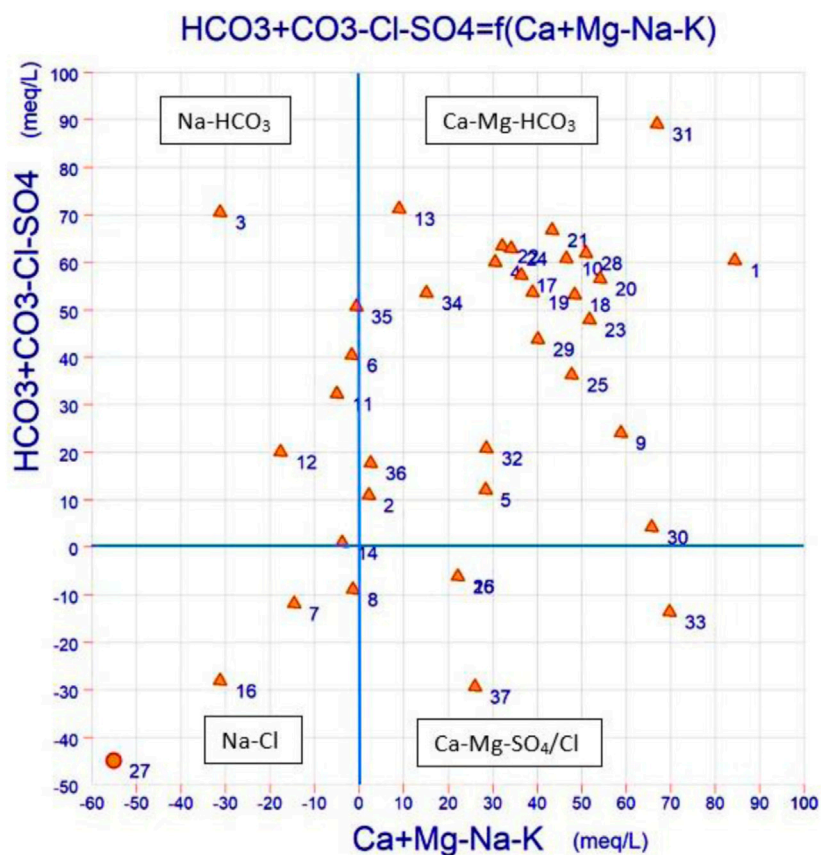
FIGURE 5
Chadha diagram showing the hydrochemical facies of the groundwater samples.

the plants root. The exchange results in two types of soils, saline soils formed when Na^+ reacted with Cl^- while the alkaline soil when Na^+ reacted with HCO_3^- in the irrigation water (Eyankware et al., 2022). In this study, $Na^+\%$ varied between 7.8% and 77.3%. On the basis of $Na^+\%$, groundwater is classified as excellent for irrigation ($Na^+\%$ less than 20%), good (20%–40%), acceptable (40%–60%), doubtful (60%–80%), and unsuitable ($Na^+\%$ greater than 80%) (Khodapanah et al., 2009). The groundwater samples are plotted in Wilcox (1948) diagram (Figure 6B). Accordingly, 64.8% of the groundwater samples are projected in the excellent class zone. This class is associated with low salinity and alkali hazard. 21.6% of the samples are classified as good for irrigation with relatively high salinity and low alkali hazard. The permissible water class included 10.8% of the groundwater samples with high alkali hazard and relatively low salinity. The groundwater in this class is mostly influenced by the rock type. Only one sample is described as doubtful for irrigation purposes, and this sample is highly influenced by salinity.

3.2.3 PI %

PI is used to show the influence of irrigation water on the soil and thus evaluate the suitability of water for irrigation. PI is based on the concentrations of the alkali ions in the groundwater. In this research, PI ranged from 35.9% to 93.5%. Generally, based on PI, groundwater is categorized into three classes as class I (PI greater

than 75%), class II (PI ranges from 25% to 75%), and class III (PI less than 25%) (Singh et al., 2020). The groundwater samples in class I and Class II are excellent and good, while in class III are unsuitable for irrigation. In this investigation, 18% of the groundwater samples are projected in class I and classified as excellent for irrigation, while 82% of the samples are classified as good for irrigation.

3.2.4 PS

PS varied from −2.4 to 3.4, with a mean value of 0.5. In general, groundwater with PS greater than 5 is considered injurious to unsuitable for irrigation, PS between 3 and 5 is regarded as good, while PS less than 3 considered excellent for agricultural purposes (Raghunath, 1987). The negative values of PS have resulted from high concentration of SO_4^{-2} compared to Cl^-. This situation is mainly due to the dissolution of sulphate minerals along the groundwater flow path. Mohammed et al. (2022b) calculated the saturation indices of groundwater in the study area and indicated that gypsum ($CaSO_4$) is precipitated in most groundwater samples. This is likely to be the cause of high SO_4^{-2}. Based on the obtained values of PS, most groundwater samples are classified as excellent to good for irrigation, with some samples exhibiting negative values of PS. Even though the PS rated the majority of the samples as satisfactory for irrigation, a thorough examination of the water quality parameters is necessary to determine the suitability for irrigation. In some cases, some ions might have greater than

TABLE 4 The observed irrigation indices of the groundwater samples used in this study.

Well no.	SAR	Na⁺ %	PI %	PS
S1	0.27	7.8	44.6	0.02
S2	3.3	48.7	68.9	−0.4
S3	4.8	65.3	93.5	−0.15
S4	1.4	34.	71.7	0.43
S5	2.5	35.6	52.5	−0.66
S6	3.7	50.5	73.5	1.51
S7	3.6	57.1	83.9	1.54
S8	3.3	50.5	71.2	1
S9	0.75	20.4	51.7	0.62
S10	1.1	26.5	55	0.13
S11	2.6	52.3	82.3	0.12
S12	3.9	58.7	85.7	1.43
S13	2.5	45.3	75.3	−0.06
S14	3.7	51.6	70.6	2.34
S15	1.9	38.6	63.5	0.55
S16	7.3	65.3	77	2.11
S17	1.4	31.6	60.1	0.06
S18	1.1	25.5	64.8	0.61
S19	1.3	30.3	62.2	0.08
S20	0.82	22.7	61.7	−0.04
S21	1.1	28.1	63.1	−0.06
S22	1.4	33.7	67.2	0.04
S23	1	23.9	62.7	−0.65
S24	1.4	32.7	66.1	−0.01
S25	1	25.8	56.3	0.1
S26	1.9	38.6	63.5	0.55
S27	13.8	77.3	84.6	−2.4
S28	0.74	24.3	67.4	0.1
S29	0.94	29.7	71	0.42
S30	0.55	16.9	60.1	3.4
S31	0.49	16.3	62	0.08
S32	1.6	35.4	63	1.23
S33	0.7	15	35.9	3.22
S34	2.6	42.2	65.7	0.47
S35	2.6	50	83.7	0.23
S36	3	48.4	71	−0.007
S37	1.4	36.8	62.5	1.91

reference values which leads to changing soil characteristics or influencing plant growth (Maia and Rodriguesda, 2012). As a result, it can be indicated that the groundwater samples with too low negative values can be considered injurious for irrigation since a high concentration of SO_4^{-2} is remarkably affecting plant growth (Papadopoulos, 1986).

In conclusion, evaluating groundwater use for agricultural purposes is comprehensively achieved by considering several combined indices. It can also be said that the use of a certain type of water depends on the type of plant and its tolerance to salinity or its sensitivity to a certain parameter. The type of soil, whether it is acidic or alkaline, is also influencing plant growth. The quantity of water that a particular plant needs for growth may be related to one index rather than the other. For example, a low permeability index reduces the amount of water that can be absorbed by the plant and thus may affect the growth of plants.

3.3 Computational intelligence results

3.3.1 Model performance in prediction of irrigation indices

In this study, the SAR, Na%, PI, and PS values were estimated using the KNN, GPR, RF, and SVR models, and the performance of each model was evaluated based on the R^2, MAE, RMSE, and testing stages, as shown in Table 5. In general, the developed models, except for SVR did not provide adequate modeling of the SAR parameter in groundwater and did not produce satisfactory SAR estimates based on the performance criteria used. However, the results for SAR estimation showed that the SVR model was the best performer, with the highest R^2 value of 0.83 and the lowest error performance values, including MAE=0.76 and RMSE = 1.45, in the validation phase (Table 5). For clarity, the prediction of SAR parameter by the best-performing model, which is the SVR, will be shown in scatter plots as seen in Figure 7. The results of the modeling for the Na% parameter in groundwater showed that the SVR model was the best performer among the models tested, with the highest R^2 value of 0.84 and the lowest error performance values including MAE = 5.16 and RMSE = 6.70 in the validation phase. The RF model was the second-best performer, with an R^2 value of 0.79 and error performance values of MAE = 4.72 and RMSE = 7.71 in the validation phase. The scatter plots of the Na% parameter prediction by the best-performing models, the SVR and RF, are shown in Figure 7. The other models tested did not provide adequate modeling of the Na% parameter and did not produce satisfactory Na% estimates based on the performance criteria used.

The results of the analysis showed that all the machine learning algorithms used in the modeling process performed below an R^2 value of 0.70 for the PI indicator in the validation set. The SVR algorithm was found to be the best performer for the PI parameter, with the lowest values for both RMSE (7.10) and MAE (5.33) in the validation phase. The results indicate that the SVR algorithm performed best in predicting the PS parameter in groundwater, with the highest R^2 value of 0.73 and the lowest error performance values, including RMSE = 0.55, in the validation phase. Other developed models did not perform well in modeling the PS

FIGURE 6
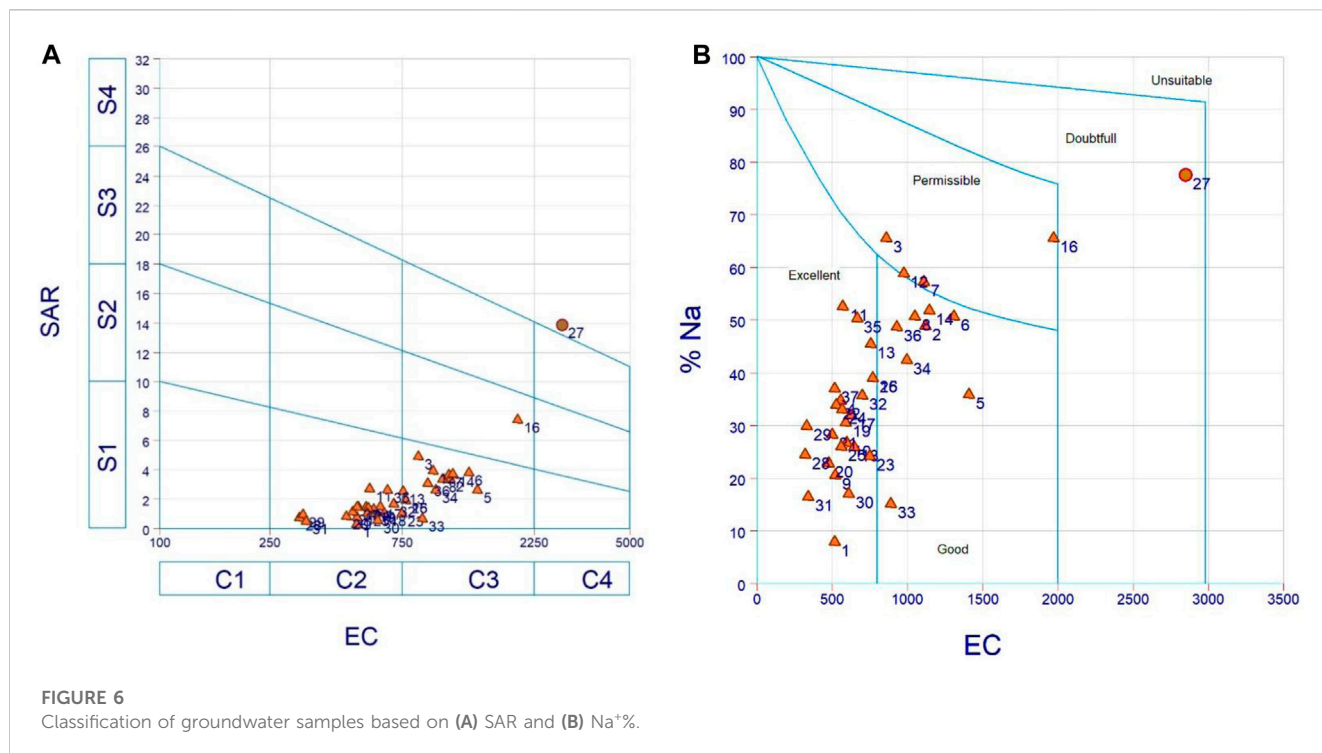Classification of groundwater samples based on **(A)** SAR and **(B)** Na+%.

TABLE 5 Machine Learning Algorithms assessments criteria of validation set for water quality parameters, root mean square error (RMSE), mean absolute error (MAE), and r squared (R^2) values.

Water quality parameters	Model	MAE	RMSE	R^2
SAR	KNN	0.86	1.82	0.73
	SVR	0.76	1.45	0.83
	GPR	1.14	2.31	0.57
	RF	0.82	2.00	0.67
Na %	KNN	7.02	9.74	0.66
	SVR	5.16	6.70	0.84
	GPR	9.01	11.33	0.54
	RF	4.72	7.71	0.79
PI	KNN	7.05	9.38	0.41
	SVR	5.33	7.10	0.66
	GPR	7.05	9.67	0.38
	RF	6.04	8.67	0.50
PS	KNN	0.43	0.85	0.37
	SVR	0.43	0.55	0.73
	GPR	0.54	0.82	0.41
	RF	0.43	0.71	0.55

parameter and did not produce satisfactory PS estimates based on the performance criteria used. The prediction of the PS parameter by the best performing SVR model will be shown in scatter plots in Figure 7. The box plots of the predicted irrigation indices using different machine learning models are shown in Figure 8. According to the one-way ANOVA analysis performed in the validation set and the results of the Tukey's test (with a significance level of $p < 0.05$), there was no statistically significant difference in the estimation of irrigation indices among the different models. However, this suggests that among the compared models, the SVR model provides the best performance in terms of the estimation of irrigation indices (Table 5 and Figure 7).

3.3.2 Spatial prediction of irrigation indices

A continuous digital map of SAR (Figure 9), Na % (Figure 10), PI (Figure 11) and PS (Figure 12) for the study area was produced from the results of the four different machine learning models using the predictor variables raster stack of the training dataset.

The spatial distribution of SAR concentrations in groundwater are predicted with all four models for the eastern Nile River region, Khartoum state, Sudan (Figure 9). In general, all four models show similar patterns for the spatial SAR distribution that only differ locally from each other. The western parts of the study area are associated with low values of SAR, likely due to the influence of groundwater recharge on groundwater samples. The highest values in the southern and central parts are generally due to the high mineralization of groundwater samples due to the dissolution of halite minerals within the Nubian formations (Abdelsalam et al., 2016; Mohammed et al., 2023b). The values of predicted SAR recorded in the KNN model ranged from 0.71 to 7.98, while in the SVR model they ranged from 0.61 to 9.40. The GPR model had a range of 1.08–6.51, and the RF model had a range of 0.56–7.42 (Figure 9). Overall, the minimum estimated SAR values were comparable among all models, except for the GPR. However, the maximum values varied among the models. The maximum uncertainty was the smallest in the RF model, ranging from
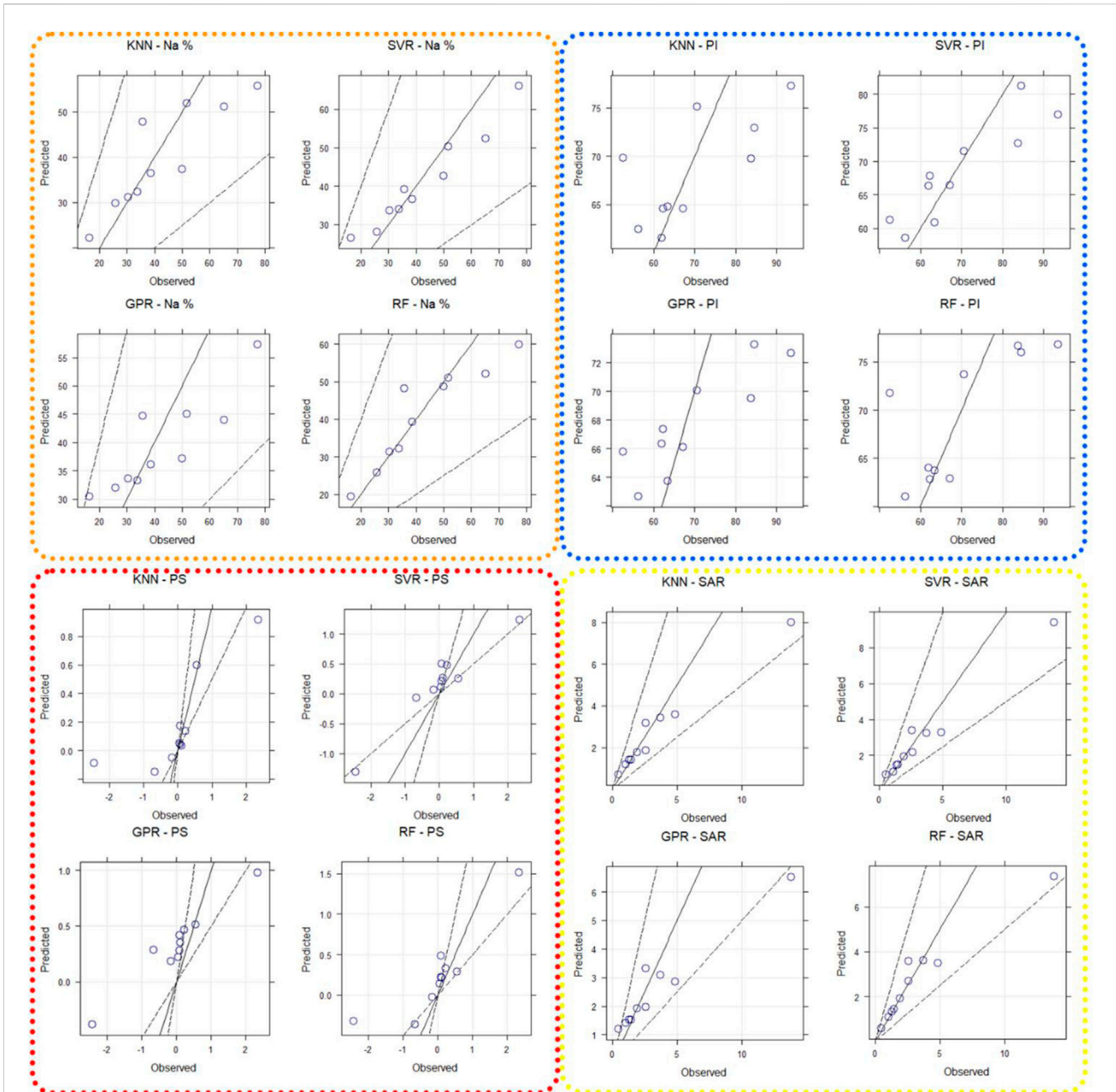
FIGURE 7
Scatter plots of observed irrigation indices and estimated irrigation indices by different machine learning models for the validation period.
Abbreviations: KNN, k-nearest neighbors; GPR, Gaussian process regression; SVR, support vector regression; RF, random forest.

0.004 to 0.31, while in all other models, the uncertainty varied from 0.01 to 3.71. Uncertainty in the SAR estimation was particularly high in areas with low sampling density, but SAR prediction uncertainty was higher in the southeast of the study area.

The results of the spatial distribution prediction for Na % using all four models in the eastern Nile River area, Khartoum state, Sudan, are presented in Figure 10. The performance of each model in predicting the Na % concentrations in groundwater can be compared based on the accuracy and uncertainty of the maps produced. The maps can be used to identify areas where the concentrations are high or low and to understand the spatial

distribution patterns of Na % in groundwater. The variation of Na^+ % shows a similar trend to that of SAR as the central and southern parts depict high concentrations relative to the rest of the study area. The results of the predicted sodium percentage (Na%) from the KNN, SVR, GPR, and RF models are presented in Figure 10. The range of Na% values predicted by the KNN model was between 22.21 and 56.83, while the SVR model had a range of 22.48–71.24. The GPR model had a range of 29.45–57.40 and the RF model had a range of 18.10–61.72. Overall, the highest estimated Na% values were similar across all models, except for the SVR model. The minimum values, however,

FIGURE 8
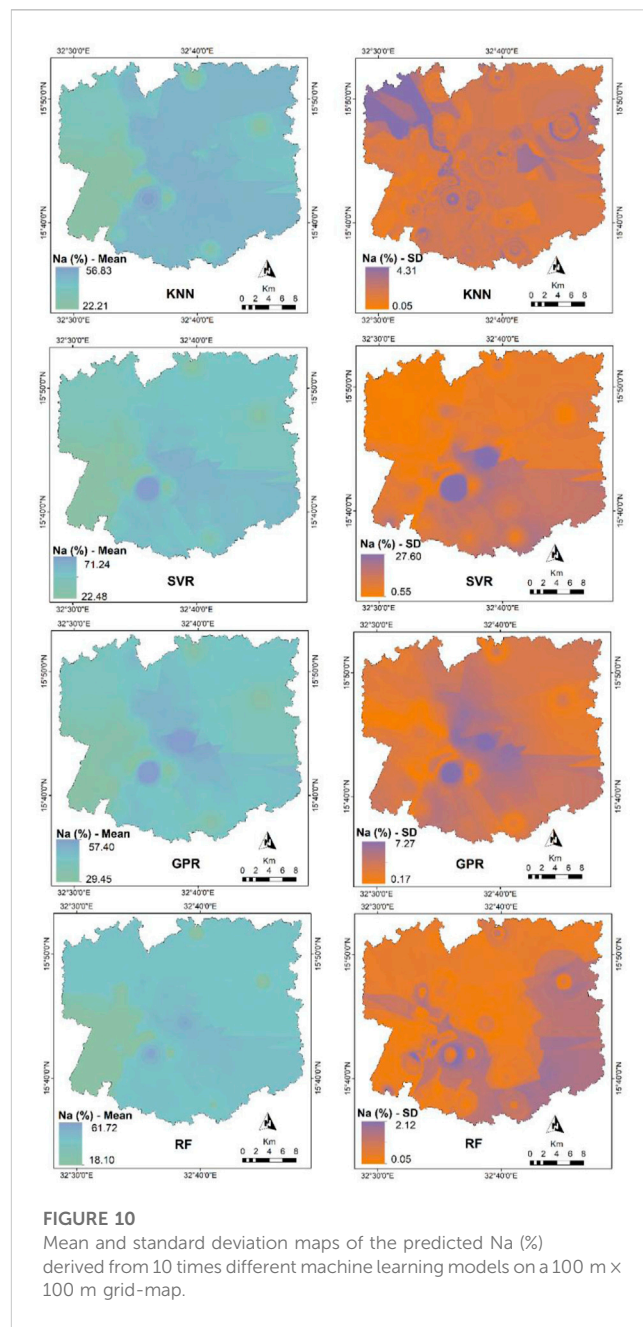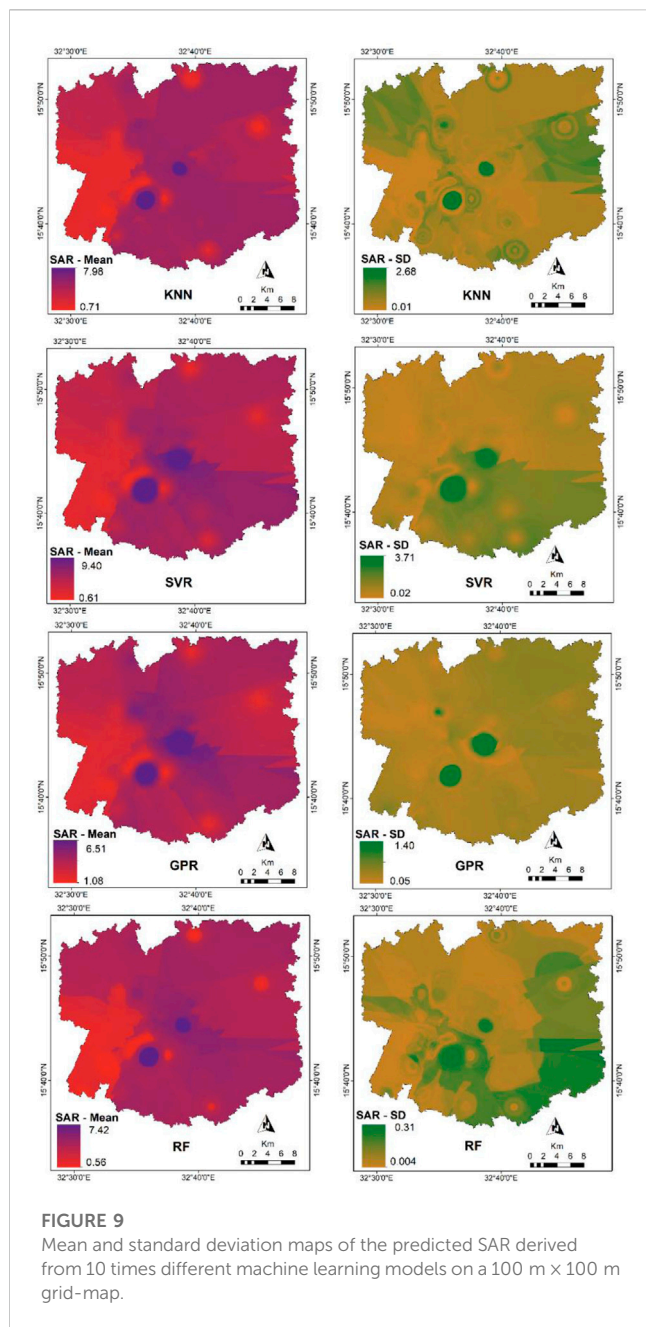Box plot showing the distribution of the predicted indices using different ML algorithms.
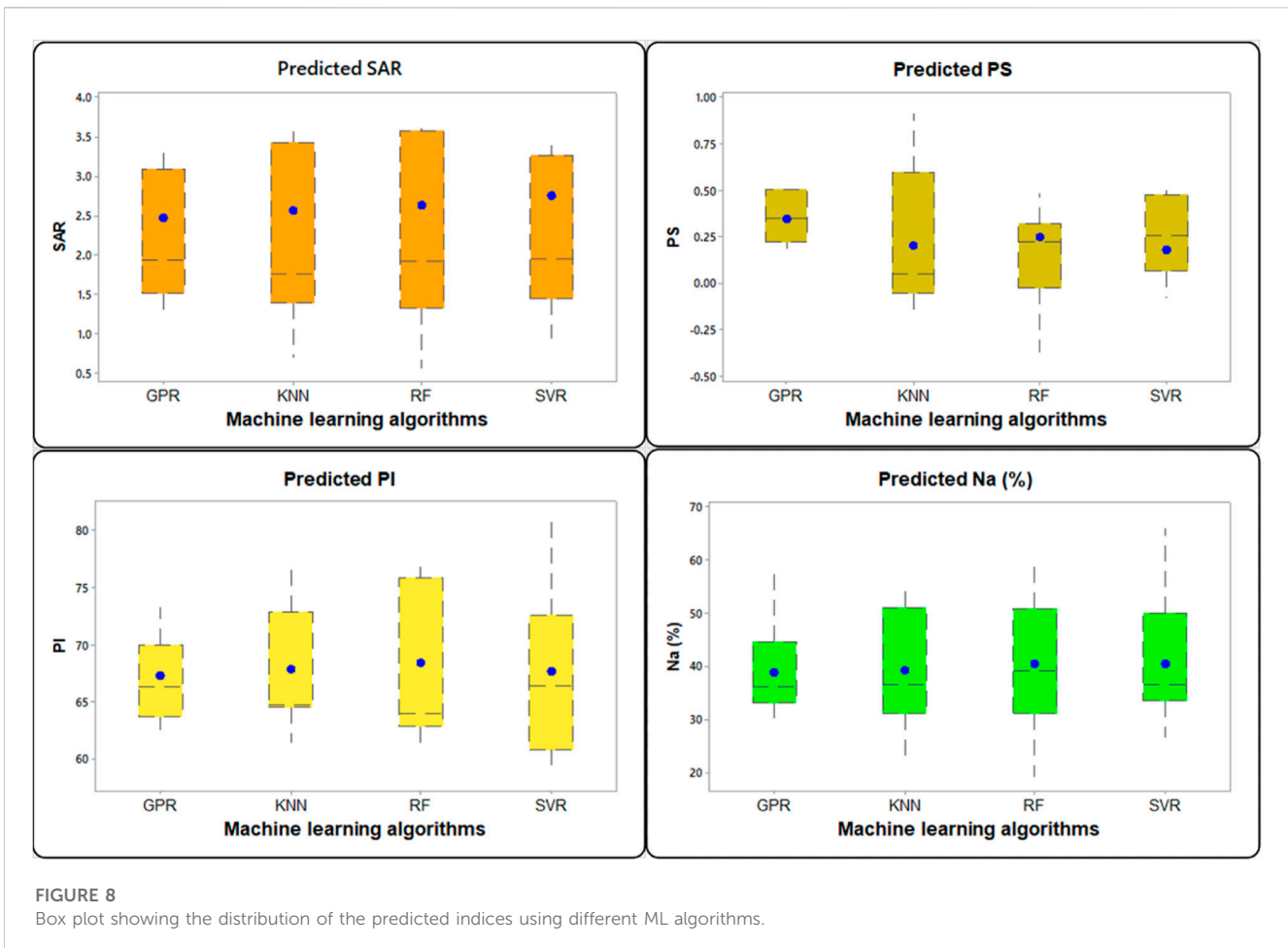
differed among the models. The RF model had the smallest uncertainty range, between 0.05 and 2.12, while in all other models, the uncertainty varied between 0.05 and 27.60. The uncertainty of Na% estimation was particularly high in areas with low sample density, with higher uncertainty observed in the southeast, east, and northwest regions of the study area.

We employed four models (KNN, SVR, GPR, and RF) to predict the spatial distribution of PI concentrations in groundwater in the Eastern Nile River area of Khartoum state, Sudan (Figure 11). Results indicated that all four models generated similar patterns of PI distribution, albeit with some local differences. The lowest PI is indicated in S33 location in the southwestern part of the study area however these samples showed low SAR and Na+%. This can be explained by two factors. The groundwater recharge from Nile River which increase the HCO_3^- concentration in the groundwater samples and therefore reduce the PI. Alternatively, it might be affected by clay nature of the quaternary deposits. The soil properties including compaction and organic contents may also influence the soil permeability (Khalaf and Hassan, 2013). In this case, the SAR and Na^+ will eventually be increased due to the ongoing ion exchange reactions. The predicted values of PI obtained from the four models ranged from 53.34 to 85.50, with the highest estimated values being recorded in the SVR model and the lowest in the RF model (Figure 11). The RF model exhibited the lowest level of uncertainty in its predictions, with a range of 0.07–2.25, while in the

other models, the uncertainty ranged from 0.06 to 14.75. The uncertainty in PI predictions was higher in areas with low sampling density and in the central and northwest regions of the study area.
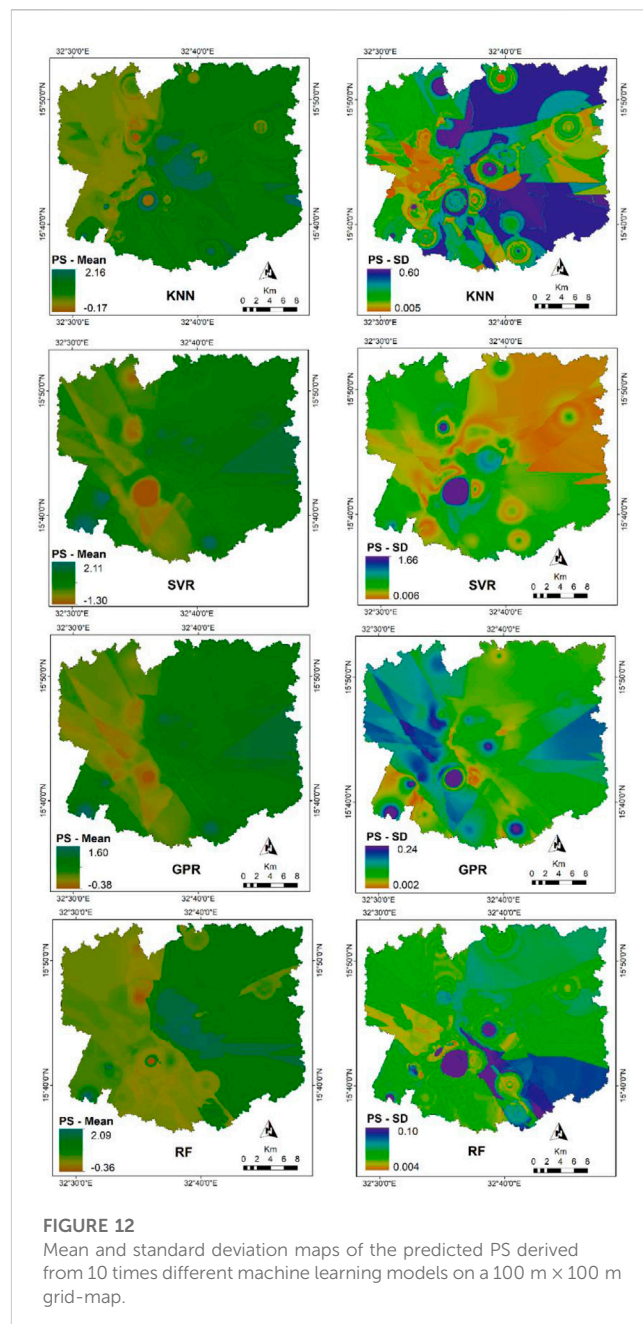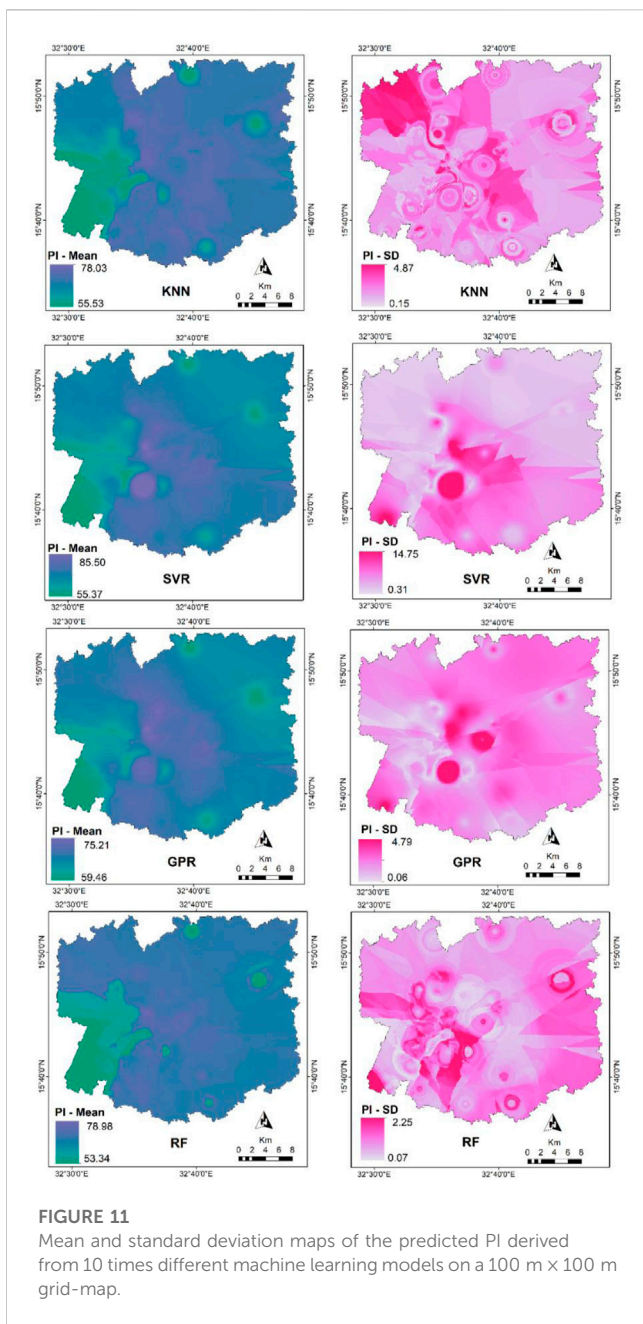
The prediction of PS concentrations in groundwater using four models (KNN, SVR, GPR, and RF) in the eastern Nile River area, Khartoum state, Sudan is presented in Figure 12. In general, all models show similar spatial patterns of PS distribution, with only minor differences in some areas. The values of predicted PS recorded by the KNN model range from −0.17 to 2.16, by the SVR model from −1.30 to 2.11, by the GPR model from −0.38 to 1.60, and by the RF model from −0.36 to 2.09. Although the maximum estimated PS values are comparable among all models, the minimum estimated values vary, with the exception of the GPR model. The maximum uncertainty was the lowest in the RF model, ranging from 0.004 to 0.10, while in all other models, the uncertainty varies from 0.006 to 1.66. Areas with low sampling density had high uncertainty in PS estimation, and areas in the central, southeast, and northeast had higher PS prediction uncertainty.

Overall findings indicate that the four models have similar capabilities in predicting the spatial distribution of irrigation indices in groundwater. However, as these models use different mathematical basis, small differences in their predictions can still exist. These differences may be due to the different weightings given to the input variables, the type of algorithm

FIGURE 9
Mean and standard deviation maps of the predicted SAR derived from 10 times different machine learning models on a 100 m × 100 m grid-map.

FIGURE 10
Mean and standard deviation maps of the predicted Na (%) derived from 10 times different machine learning models on a 100 m × 100 m grid-map.

used, and other factors. In this study, the GPR model had difficulty accurately spatially predicting the minimum and maximum values for all irrigation indices. This means that the GPR model, which is based on a normal distribution, may have trouble accurately representing data points that are significantly different from the average or have extreme values, known as outliers. This suggests that in datasets with outliers, the GPR model may not perform as well compared to other models. This suggests that the GPR model might not be suitable for modeling datasets with outliers or data that do not follow a Gaussian distribution. As supported by our results, it could be beneficial to consider other models such as RF or SVR, that can handle these types of datasets better. The findings of this study support the difficulty of the GPR model in representing the minimum and maximum values in the data set when outliers

are present. This highlights the limitations of the GPR model, especially in the context of datasets with varying data distribution and outliers, such as the groundwater data in the eastern Nile River region, Khartoum state, Sudan. Additionally, it might be worth exploring ways to transform the data to make it more Gaussian-like before using the GPR model to see if this improves the model's performance. Overall, the models can provide useful information for understanding the spatial distribution of irrigation indices in groundwater, but further validation and analysis are needed to determine the most accurate and reliable model for this purpose. This means that the overall distribution of irrigation indices in the study area, as predicted by the four models, is similar but with some slight variations in specific locations. Overall, the models produce similar results, indicating that the distribution of

FIGURE 11
Mean and standard deviation maps of the predicted PI derived from 10 times different machine learning models on a 100 m × 100 m grid-map.

FIGURE 12
Mean and standard deviation maps of the predicted PS derived from 10 times different machine learning models on a 100 m × 100 m grid-map.

irrigation indices values in the study area is relatively consistent.

In this study, an approach was presented to assess the sensitivity of the model to the available data (rasters of predicted variables) and the uncertainty of the model (Figures 9–12). Complex areas like the study area, with deltas and adjacent uplifting regions, often have strong multifactor interactions, non-linearity, and non-stationary relationships, leading to highly heterogeneous groundwater properties spatially. However, the uncertainty maps (Figures 9–12) present some conflicting results in revealing the spatial variation of irrigation indices with limited predictive variables and several water samples. To provide insight into the reliability of the irrigation indices estimates, areas with high standard deviation values should be emphasized. Other researchers have also noted the significance of uncertainty

maps in mapping irrigation indices for local areas with high groundwater heterogeneity or small sample sizes (Taşan et al., 2022). It should be acknowledged that the representation and application of uncertainty maps in machine learning predictions still require improvement.

4 Conclusion

In this research, four computational machine learning (ML) algorithms (KNN, SVR, GPR, and RF) integrated with GIS were used to evaluate the suitability of groundwater for irrigation purposes based on four irrigational indices (SAR, Na%, PI, and PS). This approach is followed to overcome the limitations of the

conventional assessment of groundwater quality parameters. Based on the modeling results, the conclusions can be summarized as follows:

- The initial analysis revealed that the groundwater samples are dominated by Ca-Mg-HCO_3 and Na-HCO_3 water types resulted from groundwater recharge and ion exchange processes.
- The observed irrigational indices indicated that the majority of the groundwater samples (60%) are excellent agricultural purposes. The remaining samples are mostly influenced by high salinity resulted from rock-water interactions.
- The GIS-based statistical approach allows the estimation of groundwater irrigation indices derived only from spatially mapped physiochemical parameters. The use of machine learning models is trained with a point data set of groundwater physiochemical parameters with exclusively spatial predictors.
- All four model types resulted in reasonable to good predictions for the spatial distribution of irrigation indices in groundwater. However, the SVR algorithm showed the best performance in predicting the irrigation indices, with the lowest RMSE value of 1.45 for SAR. The RMSE values for the other indices, Na%, PI, and PS, were 6.70, 7.10, and 0.55 in the test data set, respectively.
- The determination of the actual contributing area of the monitoring sites through a process-based representation of the groundwater flow conditions could probably improve the predictive performance of the models. However, the inclusion of locally specific groundwater flow conditions is challenging on even larger scales due to a lack of information.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

MM: Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Visualization, Writing–original draft, Writing–review and editing. FK: Formal Analysis, Investigation, Methodology, Resources, Writing–original draft. AM: Conceptualization, Data curation, Project administration, Software, Supervision, Validation, Visualization, Writing–review and editing. SA: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Writing–original draft. AA: Investigation, Methodology, Visualization, Writing–review and editing. AK: Investigation, Methodology, Project administration, Resources, Writing–original draft. NS: Conceptualization, Methodology, Resources, Software, Validation, Writing–review and editing. PS: Data curation, Investigation, Project administration, Resources, Writing–review and editing.

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdel-Fattah, M. K., Mokhtar, A., and Abdo, A. I. (2021). Application of neural network and time series modeling to study the suitability of drain water quality for irrigation: a case study from Egypt. *Environ. Sci. Pollut. Res.* 28, 898–914. doi:10.1007/s11356-020-10543-3

Abdelsalam, Y. E., Ea, E. M., and Elhadi, H. El (2016). "Problems and factors which retard the development and the utilization of groundwater for drinking purposes in the Khartoum state-Sudan," in *7th international conference on environment and engineering geophysics \& summit forum of Chinese academy of engineering on* (Dordrecht: Engineering Science and Technology), 449–451.

Ahmed, U., Mumtaz, R., Anwar, H., Shah, A. A., and Irfan, R. (2019). E ffi cient water quality prediction using supervised. *Water* 11, 1–14

Appelo, C. A. J., and Postma, D. (2005). *Geochemistry, groundwater and pollution.* 2nd. Rotterdam: Balkema.

Awad, A. Z. (1994). *Stratigraphic palyloical and paleoclogical studies in east Central Sudan (Khartoum–Kosti Basin) Late Jurassic to mid tertiary.*

Ballabio, C., Lugato, E., Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P., et al. (2019). Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. *Geoderma* 355, 113912. doi:10.1016/j.geoderma.2019.113912

Biau, G., and Scornet, E. (2016). A random forest guided tour. *Test* 25, 197–227. doi:10.1007/s11749-016-0481-7

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi:10.1023/a:1010933404324

Carroll, D. (1962). Rainwater as a chemical agent of geologic processes. *A Review-Geological Surv. water-supply* 1535, 1–16.

Chadha, D. K. (1999). A proposed new diagram for geochemical classification of natural waters and interpretation of chemical data. *Hydrogeology J.* 7, 431–439. doi:10.1007/s100400050216

Chidambaram, S., Prasanna, M. V., Venkatramanan, S., Nepolian, M., Pradeep, K., Panda, B., et al. (2022). Groundwater quality assessment for irrigation by adopting new suitability plot and spatial analysis based on fuzzy logic technique. *Environ. Res.* 204, 111729. doi:10.1016/j.envres.2021.111729

Dimple, D., Rajput, J., Al-Ansari, N., and Elbeltagi, A. (2022). Predicting irrigation water quality indices based on data-driven algorithms: case study in semiarid environment. *J. Chem.* 2022, 1–17. doi:10.1155/2022/4488446

Docheshmeh Gorgij, A., Askari, G., Taghipour, A. A., Jami, M., and Mirfardi, M. (2022). Spatiotemporal forecasting of the groundwater quality for irrigation purposes,

using deep learning method: long short-term memory (lstm). *SSRN Electron. J.* 277, 108088. doi:10.2139/ssrn.4192344

Doneen, L. D. (1964). *Notes on water quality in agriculture published as a water science and engineering*. Department of Water Sciences and Engineering, University of California. paper 4001.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., and Vapnik, V. (1996). Support vector regression machines. *Adv. neural Inf. Process. Syst.* 9.

Edition, F., and Ozolins, G. (2011). WHO guidelines for drinking-water quality. *WHO Chron.* 38, 104–108.

El Bilali, A., Taleb, A., and Brouziyne, Y. (2021). Groundwater quality forecasting using machine learning algorithms for irrigation purposes. *Agric. Water Manag.* 245, 106625. doi:10.1016/j.agwat.2020.106625

Elkrail, A. B., and Adlan, M. (2019). Groundwater flow assessment based on numerical simulation at omdurman area, Khartoum state, Sudan. *Afr. J. Geosciences* 2, 59–65.

Eyankware, M. O., Akakuru, O. C., and Eyankware, O. E. (2022). Hydrogeophysical delineation of aquifer vulnerability in parts of Nkalagu area of Abakaliki, se. Nigeria. *Sustain. Water Resour. Manag.* 8, 39–19. doi:10.1007/s40899-022-00603-6

Falowo, O. O., Akindureni, Y., and Ojo, O. (2017). Irrigation and drinking water quality index determination for groundwater quality evaluation in Akoko Northwest and Northeast areas of Ondo State, Southwestern Nigeria. *Am. J. Water Sci. Eng.* 3, 50–60. doi:10.11648/j.ajwse.20170305.11

Farah, E. A., Mustafa, E. M. A., and Kumai, H. (2000). Sources of groundwater recharge at the confluence of the Niles, Sudan. *Environ. Geol.* 39, 667–672. doi:10.1007/s002540050479

Freeze, R. A., and Cherry, J. A. (1979). *Groundwater*. Prentice-Hall.

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Tibshirani, R., and Friedman, J. (2009). "Overview of supervised learning," in *The elements of statistical learning: Data mining, inference, and prediction*, 9–41.

Hussein, M. T. (1992). On the depositional framework of the cretaceous omdurman formation in Khartoum area, Sudan. *J. Afr. Earth Sci.* 14, 559–566. doi:10.1016/0899-5362(92)90088-T

Idriss, H., Salih, I., and Sam, A. (2011). Study of radon in ground water and physicochemical parameters in Khartoum state. *J. Radioanalytical Nucl. Chem.* 290, 333–338. doi:10.1007/s10967-011-1295-4

Kelly, W. E., and Reiter, P. F. (1984). Influence of anisotropy on relations between electrical and hydraulic properties of aquifers. *J. hydrology* 74, 311–321. doi:10.1016/0022-1694(84)90021-0

Khalaf, R. M., and Hassan, W. H. (2013). Evaluation of irrigation water quality index (iwqi) for Al-dammam confined aquifer in the west and southwest of karbala city, Iraq. *Int. J. Civ. Eng. (IJCE)* 2, 21–34.

Khodapanah, L., Sulaiman, W. N. A., and Khodapanah, N. (2009). Groundwater quality assessment for different purposes in Eshtehard District, Tehran, Iran. *Eur. J. Sci. Res.* 36, 543–553.

Köhnke, M., Skala, W., and Erpenstein, K. (2017). "Nile groundwater interaction modeling in the northern Gezira plain for drought risk assessment," in *Geoscientific research in northeast africa* (CRC Press), 705–711.

Kouadra, R., and Demdoum, A. (2020). Hydrogeochemical characteristics of groundwater and quality assessment for the purposes of drinking and irrigation in Bougaa area, Northeastern Algeria. *Acta Geochim.* 39, 642–654. doi:10.1007/s11631-019-00393-3

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., et al. (2020). *caret: classification and Regression Training. R package version 6.0-86.*

Li, J., Chen, J., He, P., Chen, D., Dai, X., Jin, Q., et al. (2022). The optimal irrigation water salinity and salt component for high-yield and good-quality of tomato in Ningxia. *Agric. Water Manag.* 274, 107940. doi:10.1016/j.agwat.2022.107940

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R. news* 2, 18–22.

Liu, Z., Xu, J., Liu, M., Yin, Z., Liu, X., Yin, L., et al. (2023). Remote sensing and geostatistics in urban water-resource monitoring: a review. *Mar. Freshw. Res.* 74, 747–765.doi:10.1071/MF22167

Mohammed, M. A. A., Eltijani, A., Szabó, N. P., and Szűcs, P. (2023b). Multi-criterion approach for assessment of hydrochemical characteristics and human health risk of groundwater in Khartoum State, Sudan. *Res. Square*. doi:10.21203/rs.3.rs-2481210/v1

Maia, C. E., and Rodrigues, K. K. (2012). Proposal for an index to classify irrigation water quality: a case study in northeastern Brazil. *Rev. Bras. Ciência do Solo* 36, 823–830. da P. doi:10.1590/s0100-06832012000300013

Mallik, S., Chakraborty, A., Mishra, U., and Paul, N. (2022). Prediction of irrigation water suitability using geospatial computing approach: a case study of agartala city, India. *Environ. Sci. Pollut. Res.* doi:10.1007/s11356-022-21232-8

Masoud, M., El Osta, M., Alqarawy, A., Elsayed, S., and Gad, M. (2022). Evaluation of groundwater quality for agricultural under different conditions using water quality indices, partial least squares regression models, and GIS approaches. *Appl. Water Sci.* 12, 244–322. doi:10.1007/s13201-022-01770-9

Masoudi, R., Mousavi, S. R., Rahimabadi, P. D., Panahi, M., and Rahmani, A. (2023). Assessing data mining algorithms to predict the quality of groundwater resources for determining irrigation hazard. *Environ. Monit. Assess.* 195, 319. doi:10.1007/s10661-022-10909-9

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C., et al. (2020). *Misc functions of the department of statistics, probability theory group (formerly: E1071)*. TU Wien. Avialable at: https://cran.r-project.org/web/packages/e1071.

Mohammed, M. A. A., Khleel, N. A. A., Szabó, N. P., and Szűcs, P. (2022b). Modeling of groundwater quality index by using artificial intelligence algorithms in northern Khartoum State, Sudan. *Model. Earth Syst. Environ.* 9, 2501–2516. doi:10.1007/s40808-022-01638-6

Mohamed, A., Asmoay, A., Alarifi, S. S., and Mohammed, M. A. A. (2023). Simulation of surface and subsurface water quality in hyper-arid environments. *Hydrology* 10, 86. doi:10.3390/hydrology10040086

Mohammed, M. A. A., Abdelrahman, M. M. G., Szabó, N. P., and Szűcs, P. (2023a). Innovative hydrogeophysical approach for detecting the spatial distribution of hydraulic conductivity in bahri city, Sudan: a comparative study of csókás and heigold methods. *Sustain. Water Resour. Manag.* 4, 107–116. doi:10.1007/s40899-023-00885-4

Mohammed, M. A. A., Khleel, N. A. A., Szabó, N. P., and Szucs, P. (2022a). Development of Artificial intelligence model with aid of statistical methods for simulation of water quality indices in north Khartoum area, Sudan. *Reaserch Sq.* doi:10.21203/rs.3.rs-1679487/v1

Mohammed, M. A. A., Szabó, N. P., and Szűcs, P. (2023c). Assessment of the Nubian aquifer characteristics by combining geoelectrical and pumping test methods in the Omdurman area, Sudan. *Model. Earth Syst. Environ.* doi:10.1007/s40808-023-01767-6

Mohammed, M. A. A., Szabó, N. P., and Szűcs, P. (2023d). Characterization of groundwater aquifers using hydrogeophysical and hydrogeochemical methods in the eastern Nile River area, Khartoum State, Sudan. *Environ. Earth Sci.* 82, 219. doi:10.1007/s12665-023-10915-1

Mohammed, M. A. A., Szabó, N. P., and Szűcs, P. (2023e). Delineation of groundwater potential zones in northern Omdurman area using electrical resistivity method. *IOP Conf. Ser. Earth Environ. Sci.* 1189, 012012. doi:10.1088/1755-1315/1189/1/012012

Mohammed, M. A. A., Szabó, N. P., and Szűcs, P. (2023f). Exploring hydrogeological parameters by integration of geophysical and hydrogeological methods in northern Khartoum state, Sudan. *Groundw. Sustain. Dev.* 20, 100891. doi:10.1016/j.gsd.2022.100891

Mohammed, M. A. A., Szabó, N. P., and Szűcs, P. (2022c). Multivariate statistical and hydrochemical approaches for evaluation of groundwater quality in north Bahri city-Sudan. *Heliyon* 8, e11308. doi:10.1016/J.HELIYON.2022.E11308

Mohammed, M., Kovács, B., Szabó, N. P., and Szűcs, P. (2023g). Steady-state simulation of groundwater flow in Khartoum state, Sudan. *Pollack Period.* 1–7. doi:10.1556/606.2023.00758

Mokhtar, A., Elbeltagi, A., Gyasi-Agyei, Y., Al-Ansari, N., and Abdel-Fattah, M. K. (2022). Prediction of irrigation water quality indices based on machine learning and regression models. *Appl. Water Sci.* 12, 76. doi:10.1007/s13201-022-01590-x

Nabiollahi, K., Taghizadeh-Mehrjardi, R., Shahabi, A., Heung, B., Amirian-Chakan, A., Davari, M., et al. (2021). Assessing agricultural salt-affected land using digital soil mapping and hybridized random forests. *Geoderma* 385, 114858. doi:10.1016/j.geoderma.2020.114858

Najafzadeh, M., Homaei, F., and Mohamadi, S. (2022). Reliability evaluation of groundwater quality index using data-driven models. *Environ. Sci. Pollut. Res.* 29, 8174–8190. doi:10.1007/s11356-021-16158-6

Nasir, N., Kansal, A., Alshaltone, O., Barneih, F., Sameer, M., Shanableh, A., et al. (2022). Water quality classification using machine learning algorithms. *J. Water Process Eng.* 48, 102920. doi:10.1016/j.jwpe.2022.102920

Nong, X., Lai, C., Chen, L., Shao, D., Zhang, C., and Liang, J. (2023). Prediction modelling framework comparative analysis of dissolved oxygen concentration variations using support vector regression coupled with multiple feature engineering and optimization methods: A case study in China. *Ecol. Indic.* 146, 109845. doi:10.1016/j.ecolind.2022.109845

Nouraki, A., Alavi, M., Golabi, M., and Albaji, M. (2021). Prediction of water quality parameters using machine learning models: a case study of the karun river, Iran. *Environ. Sci. Pollut. Res.* 28, 57060–57072. doi:10.1007/s11356-021-14560-8

Osta, M. E., Masoud, M., Alqarawy, A., Elsayed, S., and Gad, M. (2022). Groundwater suitability for drinking and irrigation using water quality indices and multivariate modeling in makkah Al-mukarramah province, Saudi arabia. *WaterSwitzerl.* 14, 483. doi:10.3390/w14030483

Papadopoulos, I. (1986). Effect of high sulfate irrigation waters on soil salinity and yields 1. *Agron. J.* 78, 429–432. doi:10.2134/agronj1986.00021962007800030006x

Pipia, L., Muñoz-Marí, J., Amin, E., Belda, S., Camps-Valls, G., and Verrelst, J. (2019). Fusing optical and SAR time series for LAI gap fillingwith multioutput Gaussian processes. *Remote Sens. Environ.* 235, 111452. doi:10.1016/j.rse.2019.111452

Raghunath, H. M. (1987). *Groundwater*. New Delhi, India: Wiley.

Rahnama, E., Bazrafshan, O., and Asadollahfardi, G. (2020). Application of data-driven methods to predict the sodium adsorption rate (SAR) in different climates in Iran. *Arabian J. Geosciences* 13, 1160. doi:10.1007/s12517-020-06146-4

Rasmussen, C. E., and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Springer.

Rawat, K. S., Singh, S. K., and Gautam, S. K. (2018). Assessment of groundwater quality for irrigation use: a peninsular case study. *Appl. Water Sci.* 8, 233. doi:10.1007/s13201-018-0866-8

Richards, L. A. (1954). *Diagnosis and improvement of saline and alkali soils*. LWW.

Sattari, M. T., Farkhondeh, A., and Patrick Abraham, J. (2018). Estimation of sodium adsorption ratio indicator using data mining methods: a case study in urmia lake basin, Iran. *Environ. Sci. Pollut. Res.* 25, 4776–4786. doi:10.1007/s11356-017-0844-y

Sattari, M. T., Feizi, H., Colak, M. S., Ozturk, A., Apaydin, H., and Ozturk, F. (2020). Estimation of sodium adsorption ratio in a river with kernel-based and decision-tree models. *Environ. Monit. Assess.* 192, 575. doi:10.1007/s10661-020-08506-9

Sawyer, C. N., and McCarty, P. L. (1967). *Chemistry for sanitary engineers*.

Singh, B. (2020). Prediction of the sodium absorption ratio using data-driven models: a case study in Iran. *Geol. Ecol. Landscapes* 4, 1–10. doi:10.1080/24749508.2019.1568129

Singh, K. K., Tewari, G., and Kumar, S. (2020). Evaluation of groundwater quality for suitability of irrigation purposes: a case study in the udham singh nagar, uttarakhand. *J. Chem.* 2020, 1–15. doi:10.1155/2020/6924026

Taşan, M., Taşan, S., and Demir, Y. (2022). Estimation and uncertainty analysis of groundwater quality parameters in a coastal aquifer under seawater intrusion: a comparative study of deep learning and classic machine learning methods. *Environ. Sci. Pollut. Res.* 30, 2866–2890. doi:10.1007/s11356-022-22375-4

Team, R. D. C. (2009). *A language and environment for statistical computing*. Avialable at: http://www.R-project.org.

Trabelsi, F., and Ali, S. B. H. (2022). Exploring machine learning models in predicting irrigation groundwater quality indices for effective decision making in medjerda river basin, Tunisia. *Sustain. Switz.* 14, 2341. doi:10.3390/su14042341

Whiteman, A. J. (1971). *Geology of the Sudan republic*.

Wilcox, L. V. (1948). *The quality of water for irrigation use*.

Yahyaoui, N., Neji, M., Kallel, M., Wali, A., and Hajji, S. (2023). A comparative approach of ML algorithms to rank irrigation water quality: case of oriental-coast shallow aquifer in cap-bon, northeastern of Tunisia. *Model. Earth Syst. Environ.* 9, 3733–3746. doi:10.1007/s40808-023-01694-6

Yigini, Y., Olmedo, G. F., Reiter, S., Baritz, R., Viatkin, K., and Vargas, R. (2018). *Soil organic carbon mapping: cookbook*.

Yu, H., Wen, X., Wu, M., Sheng, D., Wu, J., and Zhao, Y. (2022). Data-based groundwater quality estimation and uncertainty analysis for irrigation agriculture. *Agric. Water Manag.* 262, 107423. doi:10.1016/j.agwat.2021.107423

Zhou, G., Wu, G., Zhou, X., Xu, C., Zhao, D., Lin, J., et al. (2023). Adaptive model for the water depth bias correction of bathymetric LiDAR point cloud data. *Int. J. Appl. Earth Obs. Geoinf.* 118, 103253. doi:10.1016/j.jag.2023.103253